



# Integrating e-infrastructures for remote climate data processing

### **Christian Pagé**

Research Engineer / Climate Research Domain **ECERFACS** CECI, Toulouse, France

Alessandro Spinuso, Maarten Plieger, Wim Som de Cerff, KNMI, Netherlands Iraklis Klampanos, Vangelis Karkaletsis, NCSR Demokritos, Greece Malcolm Atkinson, EPCC, The University of Edinburgh, UK

(i)

EGU2020-4658 https://doi.org/10.5194/egusphere-egu2020-4658















# **New Challenges for Science**

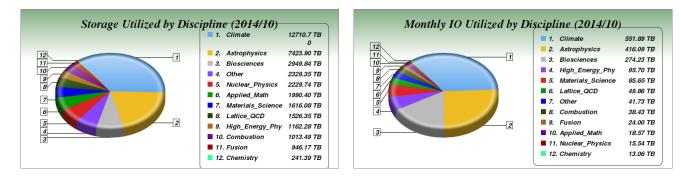
### Climate Researchers and Climate Change Community End Users

- Large needs for Storage and I/O
- Heterogeneous communities of users

### **Common Needs**

- Guidance/tools for data and scenarios subsetting: selecting a subset of representative climate scenarios
- Lower significantly the total data download size
- Calculate as much as possible remotely
- Reformat/Repackage the data into easier formats
- Access full Provenance and Lineage
- Proper Metadata description (derived data)
- Variety of Access Interfaces: GUI, OGC, REST APIs, Jupyter Notebooks, …

	CMIP5	CMIP6	CMIP7
Year	2012	2017	2022
Power factor	1	30	1000
Npp	200	357	647
Resolution [km]	100	56	31
Number of mesh points [millions]	3.2	18.1	108.4
Ensemble size	120	214	388
Number of variables	800	1068	1439
Interval of 3-dimensional output (hours)	6	4	3
Years simulated	90000	120170	161898
Storage density	0.00002	0.00002	0.00002
Distributed Archive Size (Pb)	3.19	86.05	2260.20



National Energy Research Scientific Computing Center (NERSC) Storage and I/O by Discipline



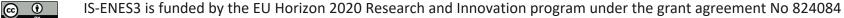


# **IS-ENES climate4impact (C4I)**

https://climate4impact.eu

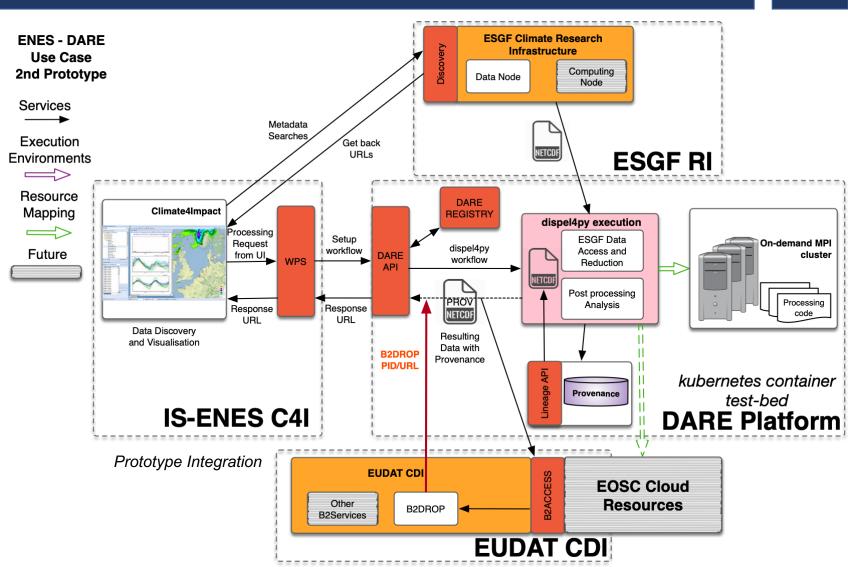
- Developed and managed by IS-ENES since 2010
- Not only UI, but also Services (WPS, WCS,..)
- Tailored for end-users
- Supports on-demand data processing
- Now containerized version
  - docker
  - docker-compose

IS-ENES INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING		g climate model da		IS-ENES   Contact   Account 🍹 🖩
Home Data disco Search Catalogs	very Downscaling Doo Explore your own catalog		Processing	٩
	r all filters		del (142) Date Geobox F	? Help
Exploring climate model data         Exploring climate model data	erature (tas) rature (tasmin) Conv. Precip	. (prc) Specific H Max. Rel. Humi Max. Rel. Min. Rel. H Rel. Humi	dity (hurs) Max. Humidity Humidity (rhsmin) E. Wir	
a rac racinged via OPADAP	Radiation Dn (rsds)         SW Radiation Dn (rsds)         SW Radiation Up (rsus)         LW Radiation Dn (rlds)         LW Radiation Up (rlus)         Diff. Radiation Dn (rsdsdiff)         Clouds (clt)	<ul> <li>Rel. Humi</li> <li>Pressure (ps)</li> <li>SL Pressure (psl)</li> <li>Pressure (pfull)</li> </ul>	dity (hur) Act. Evap. (evspsbl) Pot. Evap. (evspsblpot) Soil Evap. (evspsblsoi) Canopy Evap. (evspsblv	reg)
		CAL Econtrad So	arch and Interface	



# DARE: Execution Platform and Enabling Connections to External Resources

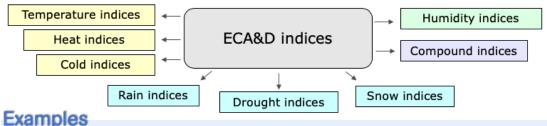
- Connect to external computing/storage resources:
  - Clouds (AWS, etc.)
  - e-infrastructures:
    - EUDAT CDI
    - European Science Cloud (EOSC)
    - DARE Platform
    - ESGF Computing Nodes (CWT)



0



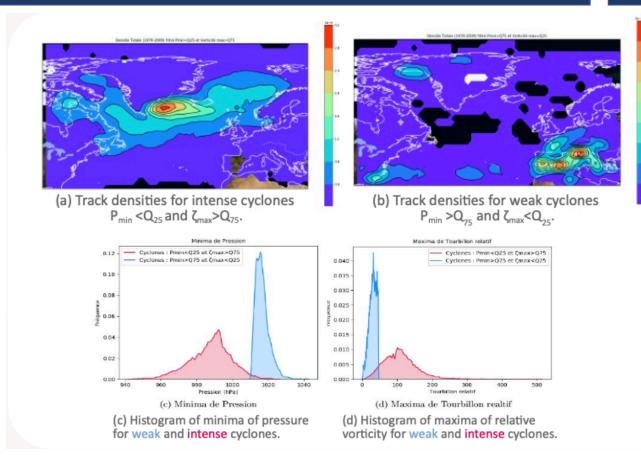
# **Leveraging Existing Complex Tools**



Intra-period extreme temperature range [° C] - ETR

- Warm days (days with mean temperature > 90th percentile of daily mean temperature) TG90p
- Summer days (days with max temperature > 25 ° C) SU

Calculating climate indices and climate indicators https://github.com/cerfacs-globc/icclim



Extra-Tropical Cyclone Tracking in Climate Simulations https://github.com/cerfacs-globc/cyclone\_tracking



# **Discussions material**

### Take Home 🟠 Messages 🜌

1. Climate Datasets grow to represent our climate more accurately

2. They require new ways of working using new data-analysis methods supported by new infrastructure

- climate scientists conduct their work through web gateways
- which exploit the power of data and computational platforms
- minimizing data movement
- avoiding the need for local resources
- widening the take up of the new data

3. DARE is a pioneering demonstration of these new methods of working on new platforms

- clouds, clusters, EOSC, EUDAT, ESGF CWT, ...
- evolving complexity hidden

 $\odot$ 

provenance and re-use as standard

4. Scientific Researchers and their developers will gain agility and productivity

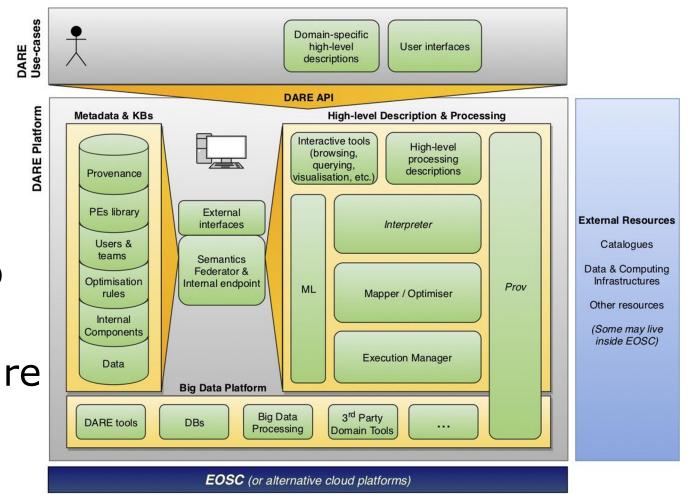
- ► <u>DARE Platform</u>
- Cyclone Tracking
- IS-ENES climate4impact

► EUDAT CDI & EOSC





- Composition of services using containers
- Across service communication using exposed REST APIs
- Scalable and flexible due to kubernetes orchestration
- Effortless cloud infrastructure deployment
- Software isolation

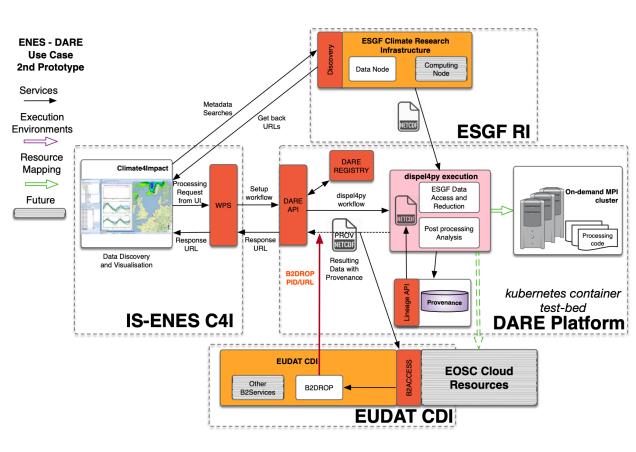






### **Simplified View of Prototype Architecture**

- ► Interfaces
  - C4I WPS / DARE API
  - DARE API / EUDAT B2DROP
  - C4I DARE Workflow / ESGF Data Nodes
  - C4I DARE Workflow / ESGF Computing Nodes
- Processing Backend
  - icclim software
- Provenance & Lineage
  - Automated by the DARE Platform
  - Community-specific information added





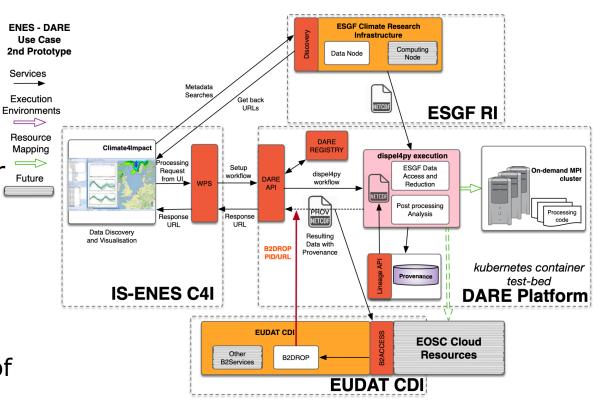
Automation/External Services to Interface and Integrate

- EUDAT-B2DROP
  - Store user output
  - > Auth B2ACCESS
- ESGF CWT
  - Delegation of processing before data transfer
  - Auth API Key
- ESGF Data Nodes
  - Retrieve Climate Datasets

### ► C4I

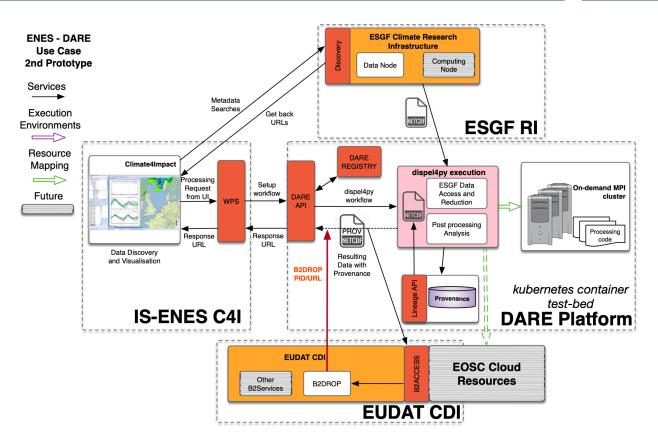
 $\odot$ 

- Use ESGF Search results to get input URLs of datasets
- Storage of output and transfer back to C4I
- Control whole workflow





- Efficient Interfacing e-Infrastructures for Researchers is Challenging
  - Technologies: fast-evolving but many are getting deprecated and obsolete
  - Authentication and Security
  - Scalability in Federated Environments
  - DARE Platform
    - Hides complexity and heterogeneity
    - Provides automated Provenance & Lineage
    - Provides EUDAT & EOSC Compatibility

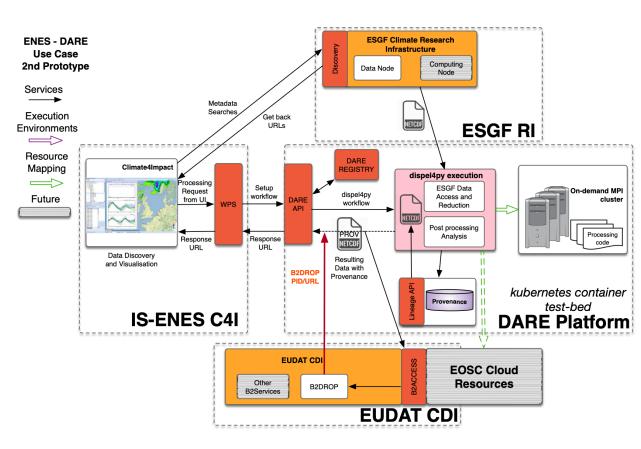


 $\odot$ 



### ► IS-ENES C4I

- User-Driven Interface Development
- Focus on Data
- Provides API
- Ease Access to large climate datasets
- EUDAT CDI
  - Solid Collection of Services for Scientific Data
  - Compatible with EOSC
  - Integrates community-specific aspects





# **Cyclone Tracking**

#### Haran Émilie<sup>1</sup>, Laloue Alice<sup>1</sup>, Pourquié Clément<sup>1</sup> supervised by Pagé Christian<sup>2</sup>, Sanchez Émilia<sup>2</sup> <sup>1</sup>IENM2, École Nationale de la Météorologie, Toulouse, France <sup>2</sup>Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS), Toulouse, France

#### Abstract

This project is focused on the study of the evolution of cyclone trajectories in a future climate. A tracking code was applied on historical climate simulations and climate projections from the PRIMAVERA project. Tools to analyze the results had to be developed : maps of cyclone track densities, maps of variability of track densities, histograms.

#### Introduction

Extra-tropical cyclones are low pressure systems in midlatitudes associated to local maxima of vorticity often leading to strong winds, extreme precipitation and high waves. These cyclones affect livelihoods and infrastructure. It is therefore important, in climate, especially in the context of climate change to be able to detect them, to follow their trajectories and their evolution, and identify areas most prone to the passage of intense cyclones. This project is part of the research carried out by the Climate Modeling and Global Change team at the CERFACS and of the European PRIMAVERA project [1]. A tool for monitoring the trajectories of extra-tropical (and tropical) cyclones was developed by M. R. Sinclair [4] and improved by the University of Quebec in Montreal (UQAM) and the Meteorological Service of Canada [2]. It was then adapted by C. Pagé at CERFACS to deal with the outputs of climate models and simulations in NetCDF format.

#### Methods

ERA-Interim atmospheric reanalysis 1979-2009 NetCDF data reference

PRIMAVERA ARPEGE CNRM-CM6-1 and CNRM-CM6-1-HR historical simulations (past climate) 1979-2009 NetCDF data

PRIMAVERA ARPEGE CNRM-CM6-1 and CNRM-CM6-1-HR climate projections (future climate) 2015-2045 NetCDF data

NetCFD data extraction and processing

#### Outputs of the tracking algorithm

A cyclone is identified as a maximum of relative vorticity at 500hPa associated to a low pressure at sea level within a radius of 150km.

Cyclone tracks are composed of at least 2 points.

1 48.326 322.031 0 956.29 12 01 01 1950 1 51.128 322.031 0 967.78 12 02 01 1950 1 53.929 320.625 0 980.20 12 03 01 1950 1 56.731 320.625 0 990.30 12 04 01 1950 2 65.136 350.156 240 1004.48 12 05 01 1950 2 65.136 12.656 0 991.28 12 06 01 1950

### UQÀM 差 CERFACS

Sinclair, M. R. et al., 2004 [4]

Calculating cyclone tracks

#### Tools Track densities and Variability

A matrix of density in which each coefficient corresponds to a spatial grid point is created. A coefficient is implemented by one each time a cyclone track crosses the grid point it is associated with.

To calculate the variability, the standard error of several matrix of density is calculated.

Analysis of the outputs of the tracking algorithm with the tools developed

#### **Results and Discussion**

#### **1. Track densities of extra-tropical cyclones in the ERA-Interim reanalysis**

The tracking algorithm was first applied on data from the ERA-Interim reanalysis [3] - on which the new

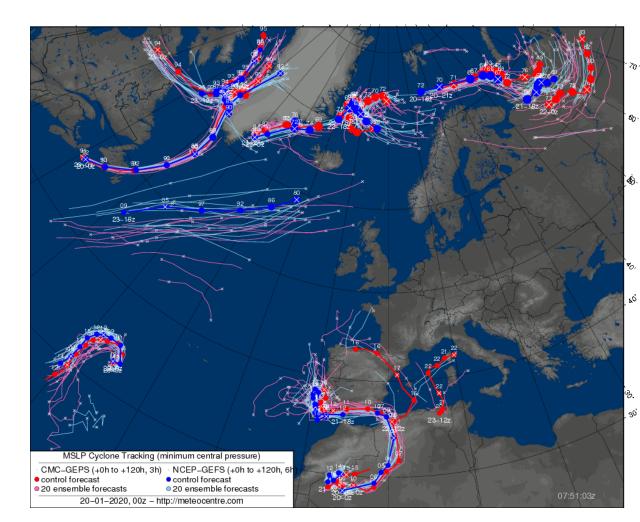
3. Track densities of extra-tropical cyclones in the climate projections (2015 – 2045)



# **ENES Use Case: Storm Tracking**

- The new ENES Use Case is about tracking storms in climate scenarios
- Initially it is for extra-tropical cyclones but the same implementation can be used for tropical cyclones







# Based on the Sinclair Methodology (1997) Group Effort

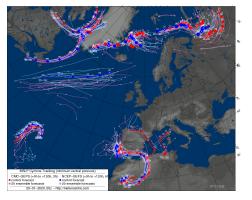


UQÀM

(00)

Canadian Meteorological Centre

Université du Québec à Montréal



High impact Lab Québec region Environment Canada

•Rabah Aider •Jean-François Caron •Louis-Philippe Caron Corina Costea Ronald Frenette Stéphane Gagnon Philippe Gachon Rares Gheti •Anne-Marie Leduc •Philippe Martin, •Milka Radojevik Christian Saad Mark R Sinclair •Katja Winger Ayrton Zadra

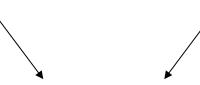
•Christian Pagé



### **Rational - Unified code**

Tropical center diagnostic simple tracking method Pressure/Vorticity centers elaborated tracking method

Several versions of the code



Tropical center and transition diagnostic

AND

Pressure/Vorticity centers

Both using same elaborated tracking method

One program, one code. Maintenance, implementation easier

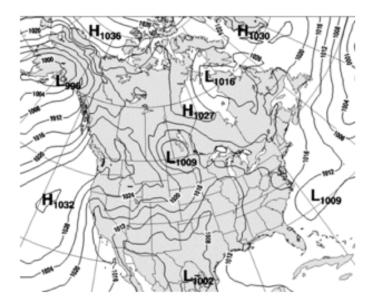


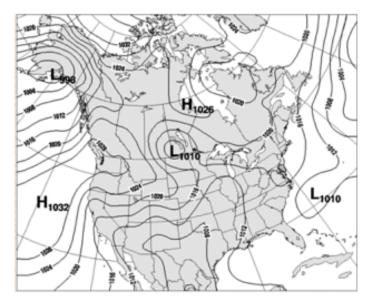
### Step 1. Preprocessing the data

For pressure center tracking

Raw NWP output

300km Cressman filter to remove small scale details







# Step 2a. Identifying minima

•Uses cubic splines between grid points for interpolation for a more precise location.

Surface pressure field unreliable over higher terrain:
 Filter with vorticy threshold.
 Threshold varies with terrain height





### Step 2b. Identifying tropical centers

- Pressure minimum
- A 300km Cressman filter is applied to the 850 hPa vorticity field
  - A vorticity max higher the 2.5x10-5 s-1 is present within 150km of pressure center.
- 850-250 hPa thickness higher then 925 dam within 150km.
  - Indicating the presence of a deep and vertical warm center
- A 10m wind speed higher then 22 kts within 225km
- A baroclinicity in the low levels
  - Measured in terms of the asymmetry of the 900-600 hPa thickness\*.

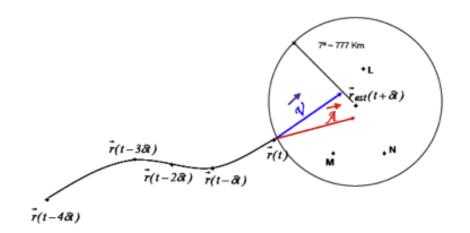
Use the Safir-Simpson surface wind thresholds to classify •Tropical Depression : V10m < 34 kts •Tropical Storm : 34 < V10m < 64 kts •Hurricane : V10m > 64 kts

\*Sinclair, M. R., 2004: Extratropical Transition of Southwest Pacific Tropical Cyclones. Part II: Midlatitude Circulation Characterisitcs, Mon. Wea. Rev., 132, p. 2149.





### **Step 3. Tracking the centers**



r(t\*) is the position of the first point of the trajectory and r(t) is the position of the current point.
wm is a weight function depending on the number of analysis per day (2,4,8,..)

$$\vec{\mathbf{r}}_{est}(\mathbf{t} + \partial \mathbf{t}) = \vec{\mathbf{r}}(\mathbf{t}) + w_m \cdot \vec{\mathcal{A}} + (1 - w_m) \cdot \vec{\mathcal{V}} \cdot \partial \mathbf{t}$$

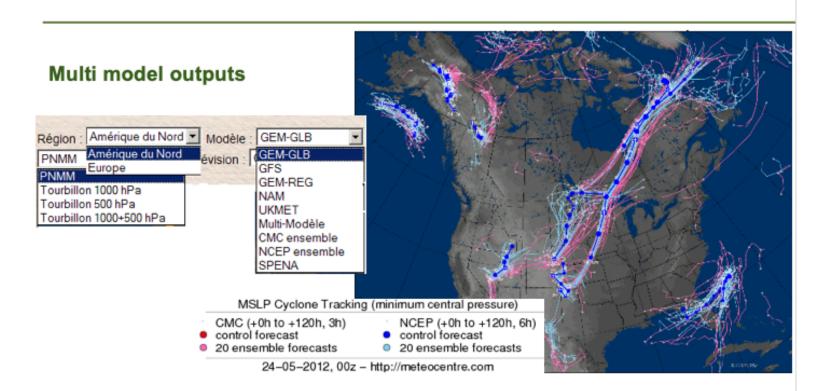
Mark R Sinclair's method

$$\vec{A} = \frac{\vec{r}(t) - \vec{r}(t - 4\partial t)}{4}$$

 $\odot$ 



### **Step 4. Outputs**



http://meteocentre.com/tracking/index e.html

From: Jean - François Caron – UK Met





### **Application - Storm impacts study**



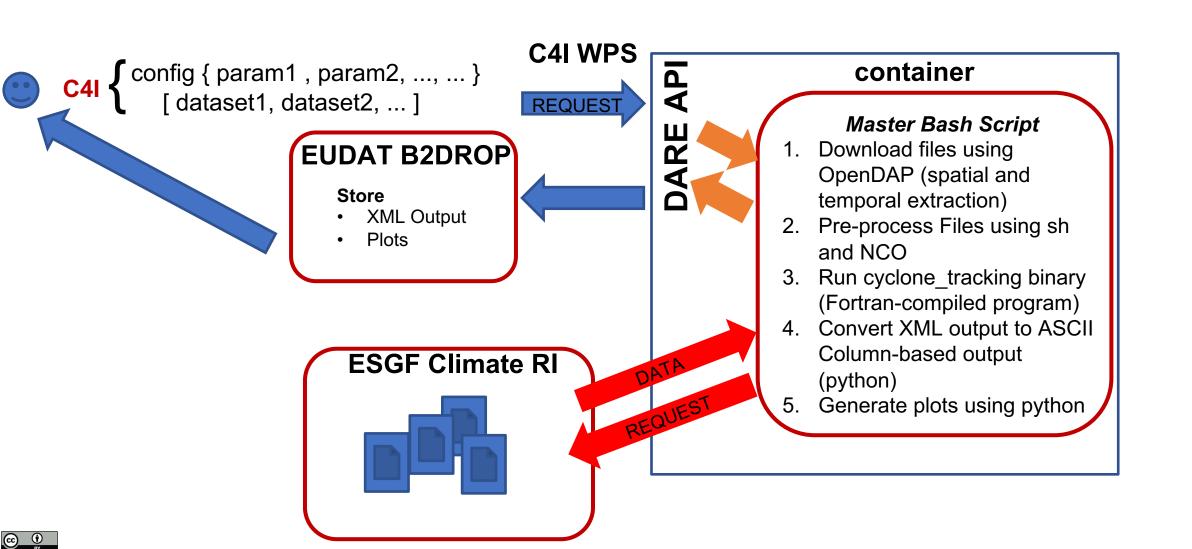
To improve our knowledge of weather systems affecting urban and surrounding zone
 Hudson Bay (MTQ project) and southern Québec
 track density, storm duration, mean circulation, mean vorticity and wind 1000 hPa

- To analyze the links between storms and hazardous events (extremes, high impact)
  2m temperatures, precipitations and 10m wind
- •To understand the impact of these extremes on population health and infrastructures.
- To predict future changes in the storm climatology and theirs effects on surface extremes
   cerca 2050





# **Storm Tracking Use Case Workflow**





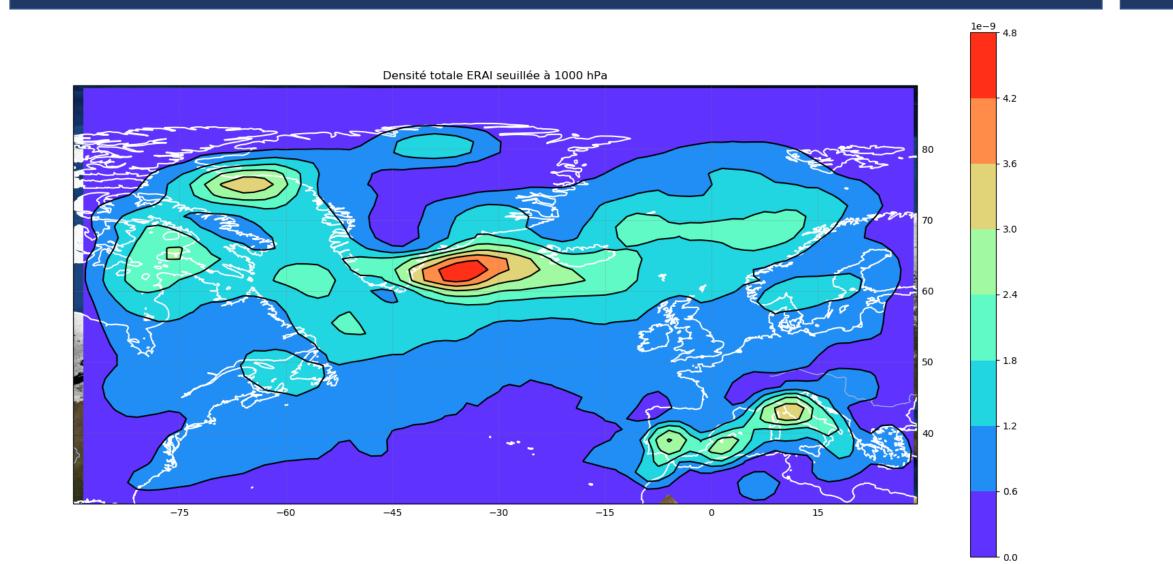
# Storm Tracking Use Case Workflow Current status

### tracking\_master.sh

- 1. Input to the script: JSON configuration file and Input data files URLs
- 2. Create directory structure
- 3. get\_files.sh : Download global sample climate scenario files using wget and hard-coded B2DROP URLs
- 4. Loop over each time period, as climate scenario files are chunked into several files over a long time period (e.g. 5-year, 10-year, etc.)
  - 5. extract\_data.sh : using NCO NetCDF tools
    - 1. Extract spatial area and time period
    - 2. Extract specific vertical levels for 3D variables (Geopotential@1000; Winds@500; ...)
    - 3. Rename coordinates and variables to standard names
    - 4. Remove extra vertical coordinates having 1 element
    - 5. Concatenate all single variables into a single file, except land-sea mask and orography
  - **6.** *make\_tracks.abs* : execute algorithm (compiled fortran) using appropriate command-line arguments :
    - 1. 3 input files: main file, land-sea mask, orography
    - 2. Output prefix filename (for XML output and NetCDF output)
    - 3. JSON Configuration file
    - 4. Variable name to use to retrieve grid parameters, dimensions and time steps
  - 7. tracking\_xml2ascii.py : Convert XML output to CSV
  - 8. Create warmstart.txt file using storm centers of last day in time period
- 6. Calculate storm density. Create several plot types using python.

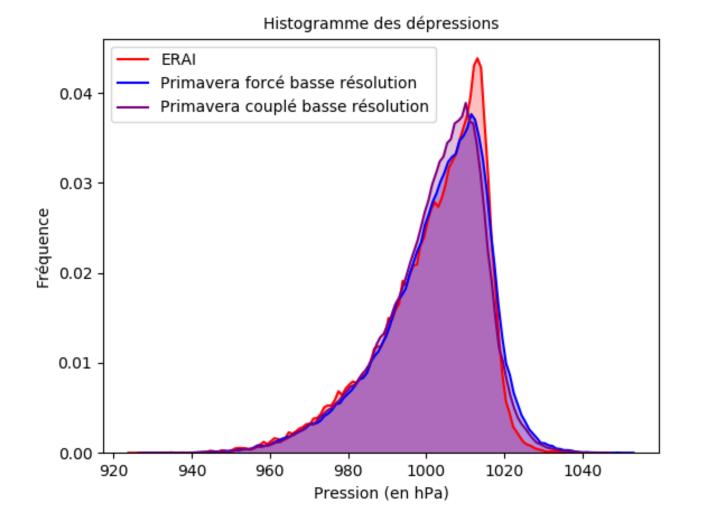


# **Storm Tracking Use Case Workflow**





# **Storm Tracking Use Case Workflow**





# **IS-ENES climate4impact**





# **Climate Data Distribution**

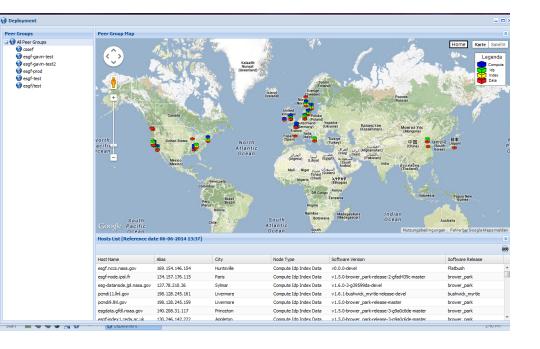
#### **ESGF Data Nodes 2019**

- 31 worldwide
- 18 in Europe (17 institutions) (coordinated by IS-ENES)









#### **IS-ENES CDI climate4impact**

- Tailored for end-users
- Supports on-demand data processing

Data Discovery Fee	edback							Search
Data Discovery	CUDUCK							
Quick filter	xtended property fi	nder Quick	filter customizer					
	0-	_		_	_			
	Tempera		👷 Precipi		💧 Humi		<del>ನ</del> ೆ Win	
	Ztas tasmin	(var) (var)	pr prc	(var) (var)	huss hurs	(var) (var)	sfcWind sfcWindmax	(var) (var)
	tasmax	(var)	oprsn	(var)	International Control of the second secon	(var)	Ouas	(var)
	ta	(var)			□rhsmin □rhs	(var) (var)	vas	(var)
					hus	(var)		
					hur	(var)		
		🕸 Radia	ation	🙆 Pre	essure	👘 Evap	oration	
		rsds	(var)	-ns	(var)	evspsbl	(var)	
	i	s-er	ies 🖌		Evoloring	alimata	model data	
		NARSTRUCTURE FOR THE EL	ROPERS NETWORK		Exploring	cimate	moueruata	
		OR DAKIN JULDI MARCO						
Data Discourse Data								
	dback	cal r1i1n1f	1 ar1 19300	101-19491	1231 nc			Search
tas_day_GFDL-	-CM4_histori					01/tas_day_GFDL-	CM4_historical_r31p1ff_prf	← to Sear
tas_day_GFDL-	-CM4_histori					101/Ras_elay_GFDL-+	_	← to Sear
tas_day_GFDL- http://expdata.gfd/noaa.g	-CM4_histori		Р/NQAA-GFDL/QFDL-C	M4/historical/r1i1p		101/tas_day_GFDL-	_	← to Sear
tas_day_GFDL-	-CM4_histori		Р/NQAA-GFDL/QFDL-C			01/tas_day_GFDL-	_	← to Sear
tas_day_GFDL- http://egdata.gfd/.noaa.g Download Dimensions	-CM4_historii		Р/NQAA-GFDL/QFDL-C	M4/historical/111	o1f1/day/tes/gr1/v201807		_	← to Sear
tas_day_GFDL- http://expdata.gl/dl.nea.g Download Dimensions • bnds of length • lat of length 18 • lat of length 18	- <b>CM4_histori</b> ev/thredds/fileServer/g 2 0 88		Р/NQAA-GFDL/QFDL-C	M4/historical/111			_	← to Sear
tas_day_GFDL- http://wsg4ta.gtdf.near.go Download Dimensions - bnds.of length 18	- <b>CM4_histori</b> ev/thredds/fileServer/g 2 0 88		Р/NQAA-GFDL/QFDL-C	M4/historical/111	o1f1/day/tes/gr1/v201807		_	← to Sear
tas_day_GFDL- http://wegdata.gtdt.nea.g Download Dimensions - binds of length - la to f length 18 - lon of length 2 - time of length 2	- <b>CM4_histori</b> ev/thredds/fileServer/g 2 0 88		Р/NQAA-GFDL/QFDL-C	M4/historical/111	o1f1/day/tes/gr1/v201807		_	- to Sear 19300101-19401231AC WMS WCS OpenDA - 40 + - 311.0
tas_day_GFDL- http://wegdata.gdf/near.go Download Dimensions - Ends of length - End frength 12 - time of length 2 - Variables	- <b>CM4_histori</b> ev/hhredda//fileServer/g 2 0 88 7300		Р/NQAA-GFDL/QFDL-C	M4/historical/111	o1f1/day/tes/gr1/v201807		_	- to Sear (19300101-19491237.nc) VMS WCS OpenDu
tas_day_GFDL- http://wegdata.gtdt.nea.g Download Dimensions - binds of length - la to f length 18 - lon of length 2 - time of length 2	- <b>CM4_histori</b> ev/hhredda//fileServer/g 2 0 88 7300		Р/NQAA-GFDL/QFDL-C	M4/historical/111	o1f1/day/tes/gr1/v201807		_	to Sear 1930001-1923271 Ar WAS WCS OpenDu # + 
tas_day_GFDL- trep:/regdata.ghtf.read.go Download Dimensions - binds of length 18 - lon of length 28 - time	CM4_historia		Р/NQAA-GFDL/QFDL-C	M4/historical/111	o1f1/day/tes/gr1/v201807		_	- to Sear (19300101-19491237.nc) VMS WCS OpenDu
tas_day_GFDL- Tag:Vergetage defease Deventions Demosions List of length List of length 2 time of length 2 double length Gouble length(1) Gouble length(1) Gouble length(1) Gouble length(1) Gouble length(1)	CM4_historie exythreadd/Westerver/g 2 0 88 7300 al () 4 (lat, bnds)		Р/NQAA-GFDL/QFDL-C	M4/historical/111	o1f1/day/tes/gr1/v201807		_	to Sear 1930001-1923271 Ar WAS WCS OpenDu # + 
tas_day_GFDL- Terr Area to the set of the s	-CM4_historia exythireddi/MieServeryp 2 0 888 7300 al () s (lat, bnds)		Р/NQAA-GFDL/QFDL-C	M4/historical/111	o1f1/day/tes/gr1/v201807		_	- 10 Sear 1930701-18439314 WKS WCS OpenDi - 4 + - 731.0 - 250 - 2610 - 2610 - 2610
tas_day_GFDL- Trap://expedite_off.exp Deutoide Deutoi	CM4_historie evolvedd./iindereer/gr 2 0 88 73300 al () s (lan, bnds) s (lan, bnds)		Р/NQAA-GFDL/QFDL-C	M4/historical/111	o1f1/day/tes/gr1/v201807		_	- to Saar S20000-14023214 VUS VOS OpenOl - 4 + - 315.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0
tas_day_GFDL- Terr Area to the set of the s	-CM4_historic 2 0 88 3300 al () 4 (lat, bnds) 4 (lat, bnds) 4 (lat, bnds)		Р/NQAA-GFDL/QFDL-C	M4/historical/111	erri (Asyntac) yr (Vy201807	UTC		- 10 Sest 1920701-18429214 WKS VCS OpenDi - 44 + - 310, - 28,0 - 46,0 - 45,0 - 45,0 - 45,0 - 45,0 - 45,0
tas_day_GFDL- Text-frequent galaxies Demtadation Dimensions List of length 1 List of length 1 List of length 1 List of length 2 Couble list (b) double list (b)	-CM4_historic 2 0 88 3300 al () 4 (lat, bnds) 4 (lat, bnds) 4 (lat, bnds)		Р/NQAA-GFDL/QFDL-C	Preview 1949-1 Nepepeere 1999	2-31 12:00:00			- to Saar S20000-14023214 VUS VOS OpenOl - 4 + - 315.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0
tas_day_GFDL- Text-frequent galaxies Demtadation Dimensions List of length 1 List of length 1 List of length 1 Could be length 1 double let Lindt double let Lindt double let Lindt double let Lindt	-CM4_historic 2 0 88 3300 al () 4 (lat, bnds) 4 (lat, bnds) 4 (lat, bnds) 4 (lat, bnds)		Р/NQAA-GFDL/QFDL-C	Preview 1949-1 Nepepeere 1999	erri (Asyntac) yr (Vy201807			- to Saar S20000-14023214 VUS VOS OpenOl - 4 + - 315.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0
tas_day_GFDL- Terretered and the constraint of t	-CM4_histori -C	frå, datarootel (CMR	лукалагацагас	Preview 1949-1 Nepepeere 1999	2-31 12:00:00			- to Saar S20000-14023214 VUS VOS OpenOl - 4 + - 315.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0
tas_day_GFDL- The restrict state states Demittable Dimensions Binds of length a last of length a last of length a itime of length a double length a do	CM4_historie withmedu/life/servery 2 0 88 2300 al () 4 (bat, bnds) at, (on) he) ds (kine, bnds) ributes for r	frå, datarootel (CMR	лукалагацагас	Preview 1949-1 Nepepeere 1999	2-31 12:00:00			- to Saar S20000-14023214 VUS VOS OpenOl - 4 + - 315.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0
tas_day_GFDL- Texture and the second	CM4_historie withmedu/life/servery 2 0 88 2300 al () 4 (bat, bnds) at, (on) he) ds (kine, bnds) ributes for r	frå, datarootel (CMR	лукалагацагас	Preview 1949-1 Nepepeere 1999	2-31 12:00:00			- to Saar S20000-14023214 VUS VOS OpenOl - 4 + - 315.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0 - 25.0



# What is the climate4impact portal?

Platform for researchers to explore climate data and perform analysis
Research data lifecycle

- Connects to ESGF web services
  - Search, Catalog Support, Security
     Several projects and experiments
- Visualization via ADAGUC Software
  - Visualization system using Web Map Services
  - Web Coverage Services for data transformation
- Analysis using (Py)WPS to perform calculations
  - icclim open-source software for on-demand climate indices calculation
  - Data sub-selection
  - Personal store for processing results.

### In production

- Deployed in the cloud
- Is one of the official CMIP6 dissemination portals

<complex-block>

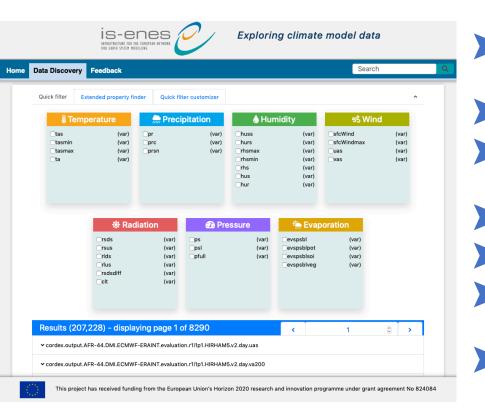
 Image: Constraint of the second s



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084



# Web based faceted search



- Drill down search results
- Tooltips for acronyms
- Quick select menus, configurable
- ES-DOC integration
  - Preview of data
  - Save Search Parameters
  - Export search list to CSV

a Discovery	Feedback								Se	earch	
Quick filter	Extended property fir	nder	Quick filt	ter customizer							^
Facet	Special Facets	Preset			Properties				Selected Properties		
var		Q		þearch		۹	•••	Search		Q	•••
variable				=wap			0	CMIP6			8
variable id				🗌 abs550ae	r		6	🕑 tas			8
variable long name			🗌 acabf			0					
	-			acabfis			•				
variant label		aclwdnt			6						
				added tre	e cover		0				
				adjunc			0				
				admeluse			0				
				adomuse			6				
				adomww			6				
Results (1	12,689) - display	ing pa	ge 1 of	508				<	1	٥	>
✓ CMIP6.CMI	P.IPSL.IPSL-CM6A-LR.1	pctCO2.r	1i1p1f1.En	non.treeFracBdlEvg	a.gr						
	P.IPSL.IPSL-CM6A-LR.1	+CO2 -	1:1p1f1 Ar	non ne dr							



III\ 🔑 🛐 🤠 🗊

in, | IS-ENES | Contact | Account | 🐂 |

add to basket

2018-10-19 11:01:582

# **Personal User Space**

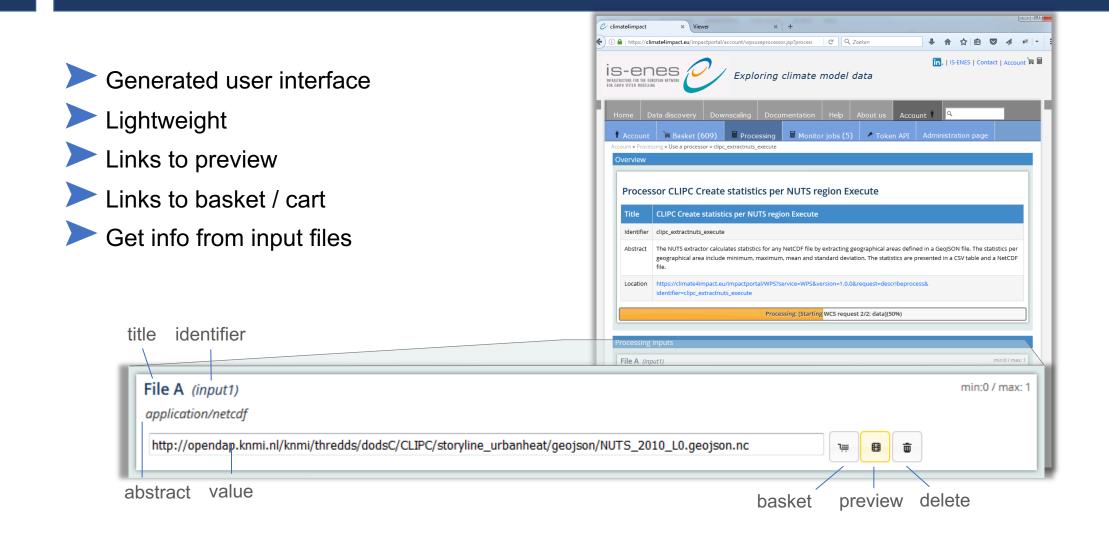
> By default the basket contains: Climate impact portal | Fo (←) → 健 🏠 🚥 🗵 🔂 🔍 Searc https://climate4impact.eu/impactportal/account/basket.jsp "Remote data" for links is-enes Exploring climate model data > "My data" for your own data 🛉 Account 🎽 Basket (163) 📓 Processing 📓 Jobs (20) 🥟 Token API 🛛 Administration pag Script based download allows to select and Account » Baske Basket download multiple files Remote data Cmip5.output1.NCC.NorESM1-ME cmip5.output1.NOAA-GFDL.GFDL-The basket allows for uploading your own tx\_0.25deg\_reg\_v17.0.nc s Amon MIROC6 historical r1i1 smax\_Amon\_MIROC6\_historical files smax dav GISS-E2-1-G hist mip5.output1.MIROC.MIROC5.h asmax day MIROC5 historical r tasmax day ACCESS1-3 historic Can be used in processing or visualization tasmax day MIROC5 rcp26 r1i1p tasmax day MIROC5 rcp26 r5i1 tasmax day MIROC5 rcp26 r1i1r NetCDF, CSV, GeoJSON, PNG 20101229120000-ESACCI-L4 GHR tasmax\_day\_MIROC5\_rcp85\_r1i1p 20101230120000-ESACCI-L4 GHF My data Share your data located in your basket with WPS\_Settings WPS Scratch tas adjust RACMO-fECEARTHothers cal 20170830po01 0000895.nd test4dz28992.nc AAIGRID\_precipitation\_2018-11-1200px-EUMETSAT\_logo.svg.png

You are logged in as https://ceda.ac.uk/openid/Maarten.Plieger

The IS-ENES project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration.

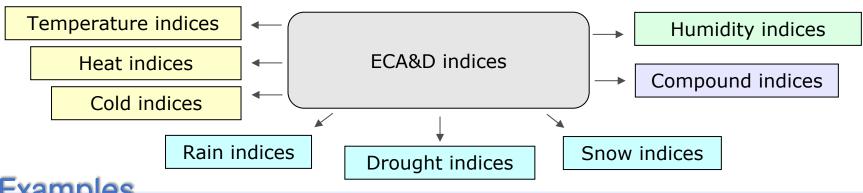


### Web processing interface for data analysis





# **Climate Indices using icclim**



### Examples

- Intra-period extreme temperature range [° C] **ETR**
- Warm days (days with mean temperature > 90th percentile of daily mean temperature) **TG90p**
- Summer days (days with max temperature > 25  $^\circ$  C) SU
- Python code developed at CERFACS, started in September 2013
  - Generic and modular approach, can be reused in other environments
  - C functions called for optimization
- I/O interface is structured for optimal performance, with wrapper functions
- Some percentile-based indices (TG10p, TX10p, TN90p, etc) using bootstrap method

icclim source code and documentation is available via https://github.com/cerfacs-globc/icclim

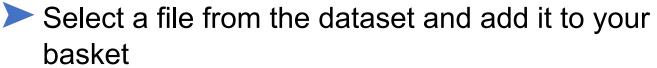
An xarray/dask fork has been done and is now at an alpha stage.



### Example: Calculating summer days (SU) 1/3

Calculate number of days where maximum temperature is above 25 degrees per European country, based on experiment RCP 2.6 and climate model MIROC5



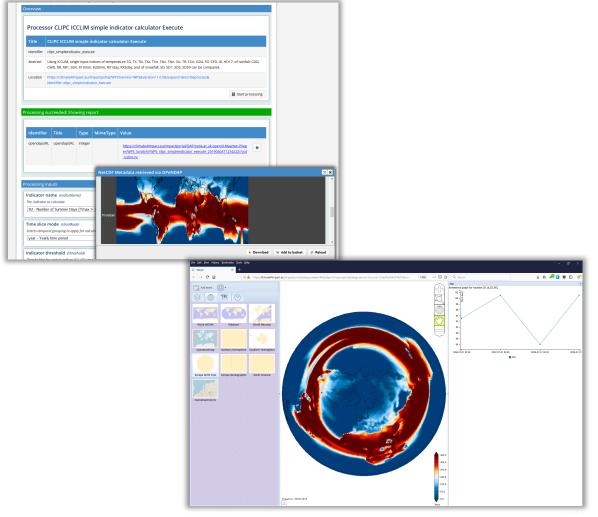






### Example: Calculating summer days (SU) 2/3

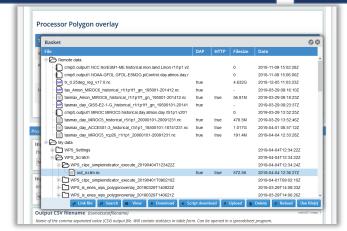
- Go to Processing and select: icclim simple indicator calculations
- Select SU, Summer days. Leave the threshold to 25 degrees Celsius
- Select the file from your basket and click "Start processing"
- Visualize the output



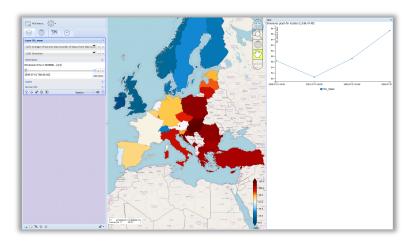


## Example: Calculating summer days (SU) 3/3

- Go to Processing and select Polygon overlay
- For "Input File B Gridded data", choose the latest result with SU from your basket. This is the most recent folder under WPS\_Scratch
- As variable select "SU", as time range select "\*"
- Click "Start processing"



**Results**: Summer days per European country for MIROC5 / RCP26 !

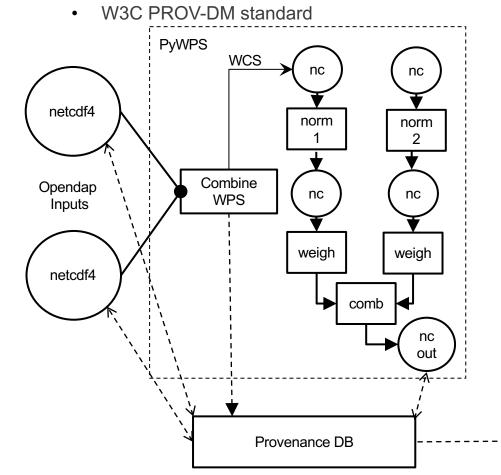




## Workflows, Provenance, Tracability and Reproducability

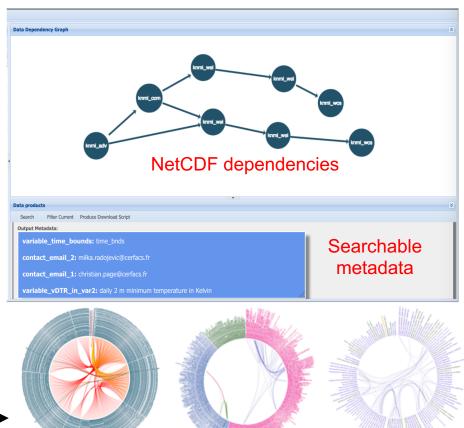
### Combine WPS, five steps involved

- Provenance module: WPS PROV
- Provenance metadata is stored in NetCDF



### Visual analytics techniques on provenance

- Highlighting data-reuse, even for cached data
- User interactions
- Exploitation of resources





# What can be improved?

#### Currently C4I handles ESGF data on file level

- > X Fragmentation of files is a barrier for many users and hurts user experience
- ➤ → Hide file structure, work with datasets and search patterns
- ► → Especially important, because now C4I is one of the official data distribution endpoints

#### > Currently the Processing services are on the same machine as the portal

- X Currently not scalable and processing load effects the portal
- ► → Make use of distributed Web Processing Services using delegation

#### Currently the frontend uses old technologies (JSP, Jquery, ExtJS)

- > X Difficult to maintain, and it is difficult to re-use results from other work
- ➤ → Migrate to ReactJS (Based on work done in the project C3S-Magic)
- $\blacktriangleright$  Good moment to re-design the user interface in collaboration with users.
- Currently the viewer is running in a separate tab
  - > By using ReactJS, it is easier to make use of an embedded viewer (adaguc-webmapjs)
- Currently provenance tracking is limited to a few processes
  - Enhance usage of W3C PROV-DM standard and WPS\_PROV toolkit
    - We are looking for users who are willing to help to improve the platform!



## **EUDAT CDI & EOSC**

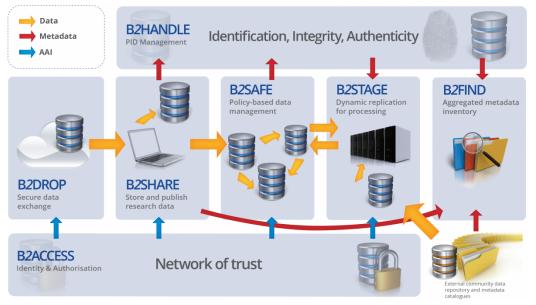




## European Landscape & Components EUDAT & EOSC

#### **EUDAT CDI B2 Service Suite**

- Integrated B2 Services
- B2ACCESS: Common AAI
- B2DROP: Secure Data Exchange
- Interface between EUDAT B2 Services and Communities infrastructures, such as Climate



### **European Open Science Cloud (EOSC)** Marketplace of Services Compute Storage Sharing, etc. FUDOPEAN OPEN About Governance Services & Resources Policy EOSC in Practice Media For providers Q SCIENCE CLOUD The European Open Science Cloud (EOSC) Offers 1.7 million European researchers and 70 million professionals in science and technology a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines. More about EOS ACCESS EOSC SERVICES & RESOURCES NETWORKIN COMPUTE STORAGE SHARING & DISCOVER







# EUDAT CDI Architecture

## Mark van de Sanden EUDAT CDI Technical Coordinator



EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065







Common Services for heterogeneous communities

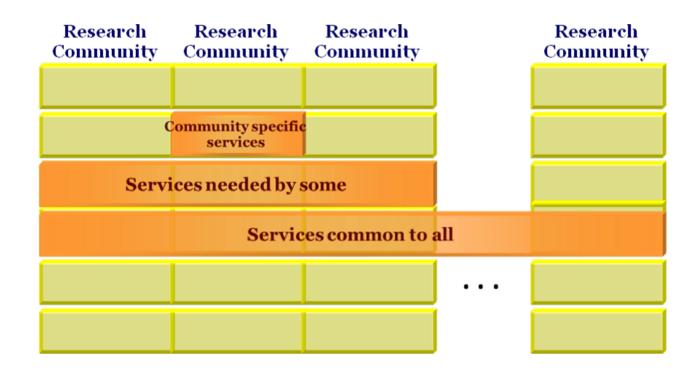
- Science data rates are exploding and will likely become continue to do so
- Building bespoke services for new communities is not cost effective
- Initial Set of Services developed as result of community needs
  - Beyond the original 'core' communities
  - New services and specific community issues highlighted







### If there are hundreds of Research Infrastructures, how many different data management systems can be sustained?



<u>www.eudat.eu</u>



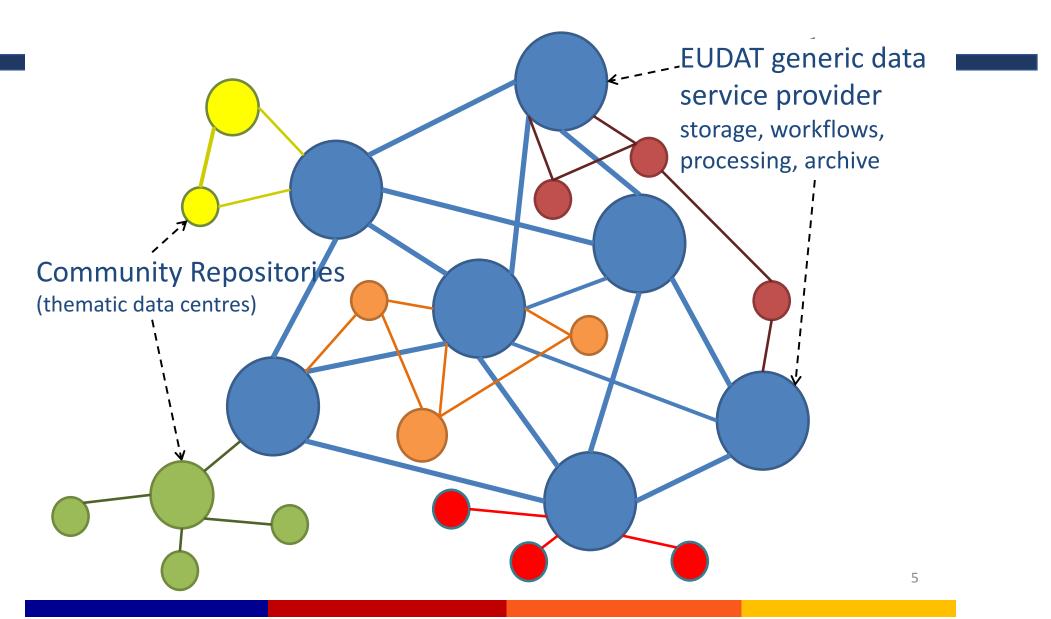


## **EUDAT CDI**

- EUDAT CDI is partner agreement
- Signed September 2016
- Total of 22 partners
- Commitment to sustain the EUDAT CDI pan-European Infrastructure for 10 years
- Partnership for Generic and Thematic service providers (level 1 and 2)
- Membership fee to become CDI partner
- No direct funding for service development
- Plan and organise participating in EC and community projects

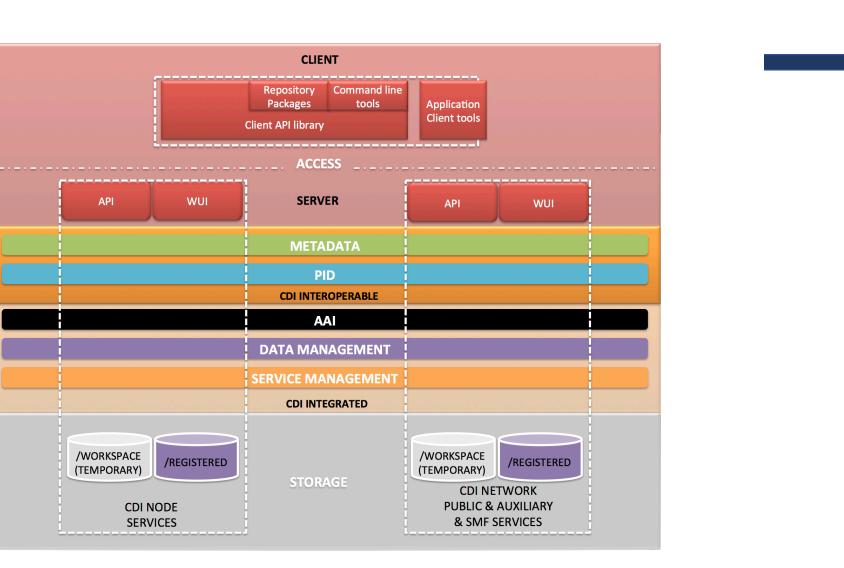
# **EUDAT** Collaborative Data Infrastructure (CDI)

ARE





## **CDI Architecture**

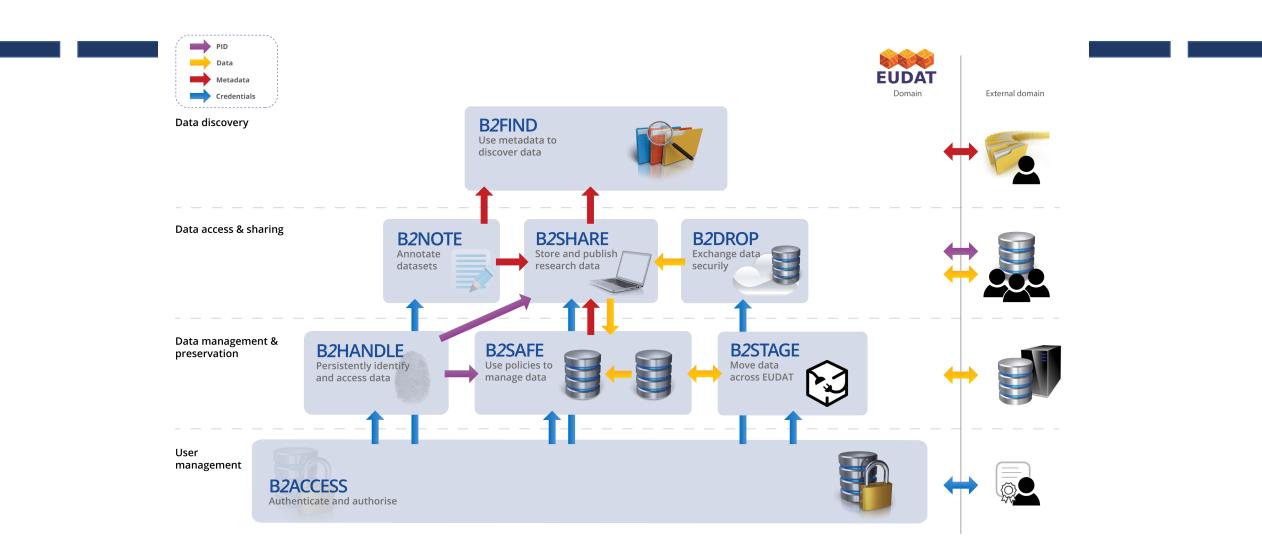


RE





### Service Diagram





### **Evolution of the CDI**

